FAQs

Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept

January 31, 2023

On January 31, 2023, the Census Bureau released a Proof of Concept to help data users understand how a new disclosure avoidance framework based on differential privacy may impact the 2020 Detailed DHC-A. The Proof of Concept includes proposed content and disclosure avoidance settings, which have not been finalized by the Census Bureau.

The Proof of Concept will help data users understand how mechanisms we used to implement differential privacy may impact the Detailed DHC-A. We'll ask for public feedback and take that into consideration as we make final decisions.

1. **What is the Proof of Concept?**
   The Proof of Concept will help data users understand how the 2020 Detailed DHC-A methodology used to protect respondent confidentiality compares to the 2010 methodology. We do this by showing published 2010 counts and comparing them to their equivalent 2010 counts with the 2020 disclosure avoidance methodology applied. The Proof of Concept will not include accompanying summary files or a complete 2010 Census data set. Rather, it will provide written documentation explaining the methodology, summary metrics, and example data. We are asking users to evaluate the data's anticipated fitness-for-use based on the proposed thresholds and margins of error.

2. **What data does the Proof of Concept use?**
   The Proof of Concept uses three types of data from the 2010 Census. The first type is the published 2010 counts. These data have swapping and partially synthetic group quarters data– the type of disclosure avoidance used in the 2010 Census – already applied to them. The second type of data is the "noise infused" counts which are 2010 Census data with differential privacy – the type of disclosure avoidance used in 2020 – applied to them. Both types of data started with the values obtained during the 2010 Census collection -- the enumerated count. Differences between those two types of counts therefore reflect the two different privacy methods used. Finally, the third type of data is summary metrics that compare the enumerated 2010 Census data to the noise infused counts, providing users with an overall assessment of the accuracy of the data in the Proof of Concept.

3. **Why does this product only include a sample of data instead of a full data set?**
   The examples included in the Proof of Concept are meant to be representative of data for detailed racial and ethnic groups and American Indian and Alaska Native tribes and villages. Unlike demonstration products for the Demographic and Housing Characteristics File (DHC), the accuracy of these data is pre-set using predetermined margins of error. These margins of error – or differences due to differential privacy mechanisms – are known in advance and applied to all detailed groups, regardless of size. The examples provided demonstrate the accuracy of data for groups of comparable population sizes.

4. **What are examples of detailed and regional groups in the Proof of Concept?**

   Detailed groups include disaggregated groups such as Puerto Rican, Chinese, the Navajo Nation, Samoan, etc.

   Regional groups include groups such as Caribbean, East Asian, American Indian, Polynesian, etc. Regional groups were not published in 2010, meaning these tabulations are new to the Detailed DHC-A.

5. **Using the settings in the Proof of Concept, what table will my detailed race or ethnic group or American Indian and Alaska Native tribe or village receive in the Detailed DHC-A?**
   The Detailed DHC-A uses an adaptive design that determines the amount of data racial and ethnic groups and American Indian and Alaska Native tribes and villages receive based on population size and geography level. By using minimum population counts (i.e., thresholds) to determine how much data are produced for each group by geography level, we are able to produce accurate data for these groups, while ensuring sufficient confidentiality protections.
   The following groups will receive total population counts *only*:
   - Detailed groups with noise-infused counts of less than 500 at the nation and state levels.
   - Detailed groups with noise-infused counts of 22-999 in counties, tracts, places and American Indian/Alaska Native/Native Hawaiian (AIANNH) areas.
   - Regional groups with noise-infused counts of less than 5,000 at the nation and state levels.
   - Regional groups with noise-infused counts of 94-4,999 in counties, tracts, and places. Regional groups are not available for AIANNH areas.

   All groups larger than those listed above are eligible for age by sex data. When detailed age by sex data are available, groups will receive one of three age category types based on the size of the group. The age by sex tables available are a four-category table, a nine-category table, and a 23-category table. No group will receive all three age category tables.

6. **How accurate are the data in the Detailed DHC-A Proof of Concept?**
   The Proof of Concept shows that the algorithm produces accurate data:

   - Over 95% of the noise infused counts for each racial and ethnic category are within their expected MOE.
   - The average difference between the enumerated count and the noise infused count for racial and ethnic groups and American Indian and Alaska Native tribes and villages at the national level ranges from +/- 0.5 to +/-1.35.
   - Less than 1% of noise infused counts are outliers for each racial and ethnic category, meaning they are outside their MOE by more than twice the value of the MOE. For example, given an MOE of 3, fewer than 1% of noise infused counts are +/-6 from their enumerated counts.

   For more information on how many noise-infused counts fall within their expected margin of error, how many outliers there are, and more about the quality of these data, please see the Metrics documentation.

7. **How is the Detailed DHC-A different from other 2020 Census data products?**

The Redistricting Data, Demographic and Housing Characteristics File (DHC), and Demographic Profile use the Top-Down Algorithm. In contrast, the Detailed DHC-A uses an algorithm called SafeTab-P. This algorithm produces noise-infused total population counts as well as noise-infused sex by age statistics for detailed racial and ethnic groups and American Indian and Alaska Native tribes and villages. This table describes the differences between these two algorithms:

Top-Down Algorithm (TDA) (Redistricting Summary File, DHC, and Demographic Profile)	**SafeTab-P (Detailed DHC-A)**
Algorithm produces privacy-protected microdata and tabulations are built from those microdata.	Algorithm directly produces privacy-protected tabulations for each geography.
Counts are produced for the nation level and counts for lower-level geographies are controlled to the national counts, so all geographies aggregate as expected (e.g., state counts sum to national count).	Counts are produced for each geography independently, and there is no requirement that geographies aggregate as expected (e.g., states will not add up to the national count).
When aggregating data, the variability of the statistical noise is controlled so that the statistics become more accurate for larger-population geographies. For example, the Census Bureau recommends data users aggregate block–level data to meaningful geographies to get more accurate data.	When aggregating data, they generally become more variable the more you aggregate. Data users should use the published statistic they are interested in, when available, rather than aggregating data. Users should use caution when aggregating data for custom geographies.
The Redistricting Data (P.L. 94-171) Summary File is used as input when processing the DHC and Demographic Profile, so these three data products are mutually consistent.	The Detailed DHC-A will not be consistent with other 2020 Census data products.
Overall accuracy and accuracy for characteristics and geographies can be targeted but the exact levels of accuracy cannot be known in advance, except in the case of the redistricting data, where the algorithm was tuned to meet pre-specified accuracy targets for total population, largest race/ethnicity group as a proportion of total population, and tract-level race and ethnicity counts.	All margins of error are determined in advance and are met 95% of the time.
Does not use adaptive design.	Uses adaptive design to determine the amount of data provided based on each racial and ethnic group's population size for a given geography.

8. How is the Detailed DHC-A different from 2010 Census data releases?

The 2020 Census Detailed DHC-A (in combination with the Detailed DHC-B) is the successor to the 2010 Census Summary File 2 (SF2) and the 2010 Census American Indian and Alaska Native

Summary File (AIANSF). The Detailed DHC-B will include household type and tenure information for the same detailed race and ethnicity groups and American Indian and Alaska Native tribal and village population groups in the Detailed DHC-A.

The Detailed DHC-A and Detailed DHC-B have fewer tables and levels of geography than the 2010 SF2 and AIANSF. To protect respondent confidentiality, we needed to reduce the amount of data released in our 2020 Census data products. Based on stakeholder feedback, we prioritized the inclusion of tables and geographies in the Detailed DHC-A and Detailed DHC-B that data users told us were critical for their work.

In previous decennial censuses, detailed data on race and ethnicity from the census were published only for a limited set of detailed Hispanic groups, detailed Asian groups, detailed Native Hawaiian and Other Pacific Islander groups, and American Indian and Alaska Native tribes and villages. Improvements allow for the tabulation of approximately 370 detailed racial and ethnic groups and 1,200 American Indian and Alaska Native tribes and villages. Data will be available for the nation, states, counties, tracts, places, and American Indian/Alaska Native/Native Hawaiian (AIANNH) areas.

Like the 2010 SF2 and the 2010 AIANSF, the Detailed DHC-A uses population thresholds to determine the amount of data available for each race and ethnic group at each level of geography. However, the thresholds used in the Proof of Concept are lower (22 for detailed groups and 94 for regional groups) than for similar 2010 data products (100 for detailed groups).

9. **Why does my group get different tables for different geographies?**
   The Detailed DHC-A uses an adaptive design that determines the amount of data racial and ethnic groups and American Indian and Alaska Native tribes and villages receive based on population size and geography level. The type of table a group qualifies for at a given geography is independent of the tables a group is qualified for at other geographic levels. Similarly, there will be differences within a given geographic level: a group that qualifies for age by sex tables in some states may only receive a total count in others due to the size of that specific group in any given state. This ensures sufficient confidentiality protection for groups of various sizes while maximizing the data provided.

10. **Will the Census Bureau provide additional guidance on how to use the Detailed DHC-A data, beyond the Proof of Concept document?**
    The information and examples provided in the Proof of Concept reflect the proposed disclosure avoidance settings and therefore this documentation is not intended to be used with the final 2020 Detailed DHC-A. The Census Bureau intends to release technical documentation with the final release of the Detailed DHC-A in August 2023.